The Best Practices in Digital Library Implementation

Jenny Cheah
7th October 2009
Kuala Lumpur Convention Centre

PROJECT PLAN

- Define Project Goal
  - Create information of value to your users

- Who is the audience?
  - Know your users’ information needs
  - Plan to reach new audiences.

- How can the audience access the digital collection?
  - To create better discovery and navigation

- Is it in line with Agency’s mission?
Staffing & Training

What are the resources needed to take on the project?

- **Staffing**
  - Setting up of Project Committee
  - Dedicated Staff to focus entirely on the project
  - Assign Staff Roles

- **Training**
  - Technical and Procedural Tasks
  - Budget for Training
  - Develop established procedures and Guidelines
  - Provide necessary Tools
  - Apply consistent quality control methods and record measures routinely

Marketing

Marketing is a Key Component of Digitization Project

- Identify Target Audience and Communicate value of collections
- Design web page that is easy to update and has good graphic design
- Create Logo or Trademark as identity for long term recognition
- Promote the new digital collections to partner organization

Cost Consideration

- Initial setup cost and on-going cost
- Preservation & Conservation
- Cataloguing/Indexing/OCR
- Type of scanning
- Database & Content Management
- Marketing Budget
- Dedicated space

Other Decisions that Affect Costing

- Text documents can be offered as images, but costs rise if you use OCR to create additional data source
- Color images cost more and require more storage space than black and white images.
- Materials that can be fed through an automatic feeder scanner cost less to digitize than those that require flatbed scanning e.g. Book Scanner
- Complex metadata and cataloguing cost more to create than simple indexing e.g. Title
Operation Model

- Operation Model Available
  - Total Outsourcing
  - Total In-House
  - Partial Outsource for example:
    - Backlog – Outsource and Day Forward – In House
    - Preparation/Conservation - In-House and Scanning & Indexing - Outsource

- Evaluate The Pros and Cons of Outsource vs In-House
- Consider Staffing & Costing

Select Collections for Digitization Project

- Evaluate intellectual value of collections
- Quantify the size of Collections
- Which collections are most appropriate for digitization?
- What is the ownership status of the materials?
- How fragile are these materials?

Developing Selection Criteria

- Material selected must meet the Goal
- Selection Criteria characterize the material and value
- The collection must be of certain theme
- Ensure comprehensive record within the theme
- Consider all the relevant material types
- Review the collections and plan digitization strategy
- Must be able to address all predictable dimensions of the theme
Preparing Collections to Digitize

- Determine all Material Types
- Availability and Fragility of the Original
- Create Inventory Listing
- Batching
- Determine Material Type e.g. physical, microfilm, photo etc
- Consider the Technical and Logistical aspects of digitization

Rights & Permissions To Publish Digital Collections

- Libraries are responsible for obtaining rights to publish and permissions to use or distribute material.
- A routine established for documentation must be in place
- Assign dedicated staff to be familiar with the Copyright Act in Malaysia.
- Post a statement to guide use of materials on your digital web site.

Preparation of Material

- Unique, damaged or fragile material – rare books, manuscripts
  - Conservation procedure and restoration prior to scanning if needed
  - Plan out Scanning Strategy to minimize turning of pages
- Different preparation for different material

Indexing & Metadata for Digital Objects

- Metadata of Digital Objects consists of:
  - Administrative Metadata
  - Structural metadata
  - Content Metadata
- Indexing standards – Title/Analysis of content/Cataloguing
- Index Placement – external database
- Indexing Methodology – Manual or OCR
TECHNOLOGY

The Best Practices In Digital Library Implementation

What machine and scanning technique to use?
Which standards and quality control methods to use?
What image resolutions levels to use?
What is the file format/naming conventions to use?
How to store the digital images and what is the best delivery method for access?

Depending on material type and condition of material.

- Book Scanner: Microbox Book2Net
- Document Scanner: Kodak Document Scanners
- Microfilm Scanners: Mekel High Performance Scanner
  Sunrise Speedscan

Capture Components

Technology Considerations

Scanning Process & Image Quality

What is the purpose and the ultimate display objective of the digitization?

- Scanning Process
  - Preparation Work varies depending on scanning methodology
  - Onsite vs offsite scanning

- Image Quality Assurance
  - Measure completeness, accuracy and clarity of an image
  - Subjective depending on the goals of the project
  - Use of Image Enhancement Software during image processing
Scanning Resolutions & File Format

Scanning resolution is dependent on the goals of the project

- Scanning Resolution
  - To create Digital Master (Heritage Standard) – 600dpi color
  - Text Image for internet viewing – 200 dpi bitonal
  - To perform full text OCR to aid in searching – 300 dpi bitonal

- File Format
  - Must conform to Industry Standards to minimize future migration
  - Digital Master – uncompressed Tiff format
  - Standard Archival – Compressed Tiff, JPEG, PDF, DJVU

Color Management & Delivery Method

- Color Management
  - To control capture and reproduction of color to have the least deviation
  - International Color Consortium (ICC) – data interchange standard

- Delivery Method of Digital Image
  - via Internet/client-server system for walk in customer
  - Use system that is non proprietary
  - Ensure Speedy loading of images
  - Thumbnail images hypertext linked to source images

Derivative Image

- Derivative Image is a stripped down version of Digital Master
- Created by using software to reduce resolution of Digital Master
- It is created for faster loading, “thumbnail” preview, improve OCR accuracy.

OCR and Manual Indexing

- OCR
  - OCR’s accuracy depends on image contrast, noise level and Font type and size
  - OCR is labor intensive if the purpose is to create full-text reproduction of the words on the source document

- Manual Indexing
  - Need to select indexing standard and format e.g. USMARC cataloging and the Dublin Core metadata schema.
  - Capture keyword from content to aid searching by users.
  - Indexing increase researchers awareness of the digital collection you are offering
PLANNING CHECKLIST

Kodak The Best Practices In Digital Library Implementation

Planning Checklist

- Define Goal and Milestones
- Define timeline, workflow chart, and evaluation method.
- Resource requirement including time for staff training
- Select Collection and segregate materials into batches for conversion and quality control.
- Include proper documentation to set as for the digital conversion methods.
- Anticipate future use and migration of the files to new delivery systems.
- Define Cataloguing/Indexing/Metadata standards
- Evaluate Operation Model
- Project Budget

Kodak – How can we help you in your project?

- Digitization Outsourcing Service
- Provides Hardware & Software required to handle various material
- Consultation Service
- Facility Setup and Management Service
- Project Management Service

ABOUT KODAK

Kodak The Best Practices In Digital Library Implementation
High Performance Technology

- State-of-the-art Imaging Technology – Loose Sheets, Bound Books, Microfilm Rolls, Microfiche, Photograph, Photo Negatives, slides etc.

- Automated Data Recognition and Content Processing – ICR/OCR/OMR

- Key Performance Index Measurement Tool

Our References

- DiGi
- Standard Chartered
- Shell
- MOL
- Mitsui O.S.K. Lines
- TM
- Malaysia Aviation
- Shell
- UNIVESITI MALAYA
- F&N
- MOX
- ANM
- express
- HSBC
- B BRAUN
- TANJUN BIN POWER SDN, BHD
- Scotiabank
- NLB
- ING
- Bank of America
- Kim Kiu Holdings Sdn Bhd

Thank You

Great Company...Great People